Integration of Heterogeneous Cloud Storages through an Intermediate WCF Service

Parvinder Kaur
Student, Department of Information Technology, Chandigarh Engineering College Landran, India
Email: pbrar67@yahoo.com

Manish Mahajan
Associate Professor, Department of Information Technology, Chandigarh Engineering College Landran, India
Email: cec.manish@gmail.com

Abstract—Many organizations provide computing resources or services to the client like hardware, software, storage etc. Storage is one of important service that is popular and commonly used today and it provides storage space online to store the data. Service providers give a storage space to consumers on the basis of renting criteria with a trust of high and powerful security. Consumer and organizations uses these storages to secure own personal data. There are many reasons come across that causes to move data from one place to another. Some of them are: reconstruction of storage, backup, new technologies, security etc. So, data migration, data transfer between different cloud storages, often requires. Achieving mechanism of Data migration between clouds is very difficult because of its different architecture to store data. For example Window Azure cloud storage creates a primary key name ID automatically whenever a new table is created. Architecture of cloud storage is depending upon its implementation done by corresponding cloud vendors (like Microsoft).

The main vision of this research work is to provide an intermediate service to move the data between both homogeneous and heterogeneous cloud storages. This intermediate service is a two way strategy that built in a SOA technology called WCF that helps the clients, organizations to move data between same and different cloud storages (Window Azure and Go Daddy cloud storages mainly used in this research work).

Index Terms—Cloud Computing, Cloud Storage, Data Migration, Service-Oriented Architecture (SOA), Web Services (WS), Windows Communication Foundation (WCF).

I. INTRODUCTION

Cloud computing like an internet where all the organizations provide the computing resources to the consumers to complete their daily routine purposes. Cloud computing has three types of services as:

- Infrastructure-as-a-service: includes services like hardware, software, storage etc.

- Platform-as-a-service: like operating system, database, and a web server and
- Software-as-a-service: that provides software to the consumers with less storage space for rent.

These services are deployed in one of the four deployment models (public cloud, private cloud, hybrid cloud and community cloud) of the cloud computing [4]. Cloud storage is an online data centric where the data is stored as hosted by the enterprises own or by third party. Cloud storage is a set of virtual servers and it is on rented as needed. Rapid remand of the cloud storage increases the growth of cloud computing and market demand becoming stronger [9].

Data migration means transferring the data between cloud storages. Secure data migration architecture must maintain the ACID (Atomicity, Consistency, Isolation and Durability) property of the data that is to be transferred.

Service oriented architecture (SOA) is based on the service oriented programming. It is developed to provide communicate between the different applications that built in different platform (drawback of object oriented programming). SOA is an emerging architectural style that builds the services and theses services are able to connect the cloud storage due to its interoperable feature. SOA provides communicate between services via messages as shown in Fig 1.
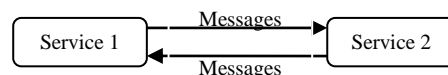

Fig. 1. SOA process.

It provides various platforms, two of them to build services as:

- Web Service
- Windows Communication Foundation.

A web service provides services that support the machine-to-machine interoperability. & WCF is a

enhance version of web services with the extension of security, velocity features.

End point of WCF: ABC (Address, Binding and Contracts).

- Here, Address specifies the location of the service like http://MyFile/Service. Client uses this location to communicate with the service,
- Binding specifies how two links will communicate. This provides various transport, encoding and protocols

Fig 2 shows a graph that compares the web services with WCF according to their velocity. If the client requested for 10 variables, then the result is received in 15ms with WCF and 156 ms with WS. This shows the reliability and fastness feature of the applications that use WCF services. When a client requests for 100 variables, then the value got in 234 ms with WCF and 1953 ms with WS [5].
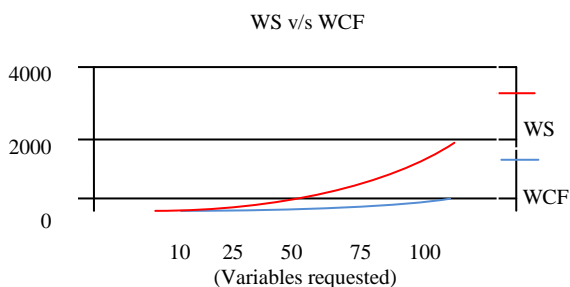
WS v/s WCF

Fig. 2. Graph showing the performance of WCF v/s Web Services.

II. RELATED WORK

A tool called Migratool, which aims to migrate the geospatial data among both distributed and heterogeneous data sources. This tool is based on three tiers architecture and it has been implemented using the J2EE architecture. It provides the interoperability environment between spatial databases through a standard structure called GML (Geography Markup Language) to represent the spatial and non-spatial data. GML is based on the XML (Extensible Markup Language) [2]. Another intermediate called intermediate Canonical Data Model (CDM) that moves the Relational Database (RDB) to the object-based/XML (Extensible Markup Language) using a predefined database standard.[1].

There are two types of migration principles used in data migration tools as: big bang migrations and trickle migrations. Big bang migration migrate the whole data into a small defined processing window. This migration looks simple but this causes the system downtime while the data is fetched from source system and loaded into the target system. Trickle migration transfer the small data between system and condition is that both systems run parallel. This process provides zero downtime but it is complex in design and requires both systems run parallel. Oracle provides a data migration strategy with advantages

of data quality that includes the extraction of the data, cleaning the data and transforms it into target system with the lacking of planning, high risks, over budget, sometime strategy fail completely and takes a lot of time [8].

Nasuni takes the three clouds Amazon S3, Microsoft Windows Azure and Rackspace and transfer the data between them with secure connection (https). It uses 40 Machines to migrate the 12TB data. 12 TB data is divides between these machines on the basis of round robin strategy. Nasuni said that data migration speed depends upon read performance of source cloud storage, write performance of target cloud storage and also based on the data migration technique that used between them with high internet rate. It estimate that it takes 40 hours to move a 12 TB volume from S3 to Azure, but only 4 hours to move from Azure to S3. Problems with this methodology are:

- Not feasible (High Cost Rate)
- High Error Rate
- Sometimes whole data migration process fails [7].

Many organizations provides data migration facility in own same cloud storage For example, to migrate data from one SQL Azure database to another SQL Azure database, you can use SQL Azure database copy and SQL Azure Data Sync Commands.

This paper focused on the data migration between any organization's cloud storage that is located at different geographical locations with different data formats. Problems associated with this are:

- Need to know API of each server
- Security Implemented on API that may helps to achieve high level of data security, accuracy and privacy
- Its working is still under development [3].

Still cloud interoperability is a big issue as no one; neither Microsoft nor Amazon supports the Unified Cloud Interface (UCI) Project proposed by the Cloud Computing Interoperability Forum (CCIF) [11].

There are many latest technologies that are used in software companies. Latest technologies include Service Oriented Architecture (SOA), Windows Presentation Foundation (WPF), Windows communication foundation (WCF), WCF provide the communication layer which act as Middleware (RPC, CORBA). WCF is a combination of Web Service, Remoting, MSMQ and COM+ [10].

There is no need to maintain the data transfer protocols or encryption in WCF as it is already defined in the configuration file only need to configure it into the configuration file and Interoperability is the fundamental characteristics of WCF because it is written into the XML file that is used to transfer the data with fast speed.

Main Advantage of using WCF technology to migrate the data between cloud storages as:

- It is based on the Trickle migration principle means both cloud storages run parallel. So this process provides no downtime while data is fetched from the source cloud storage and transferred it to the target cloud storage.
- It is feasible (negligible cost rate)
- Accurate, Secure
- Platform independent
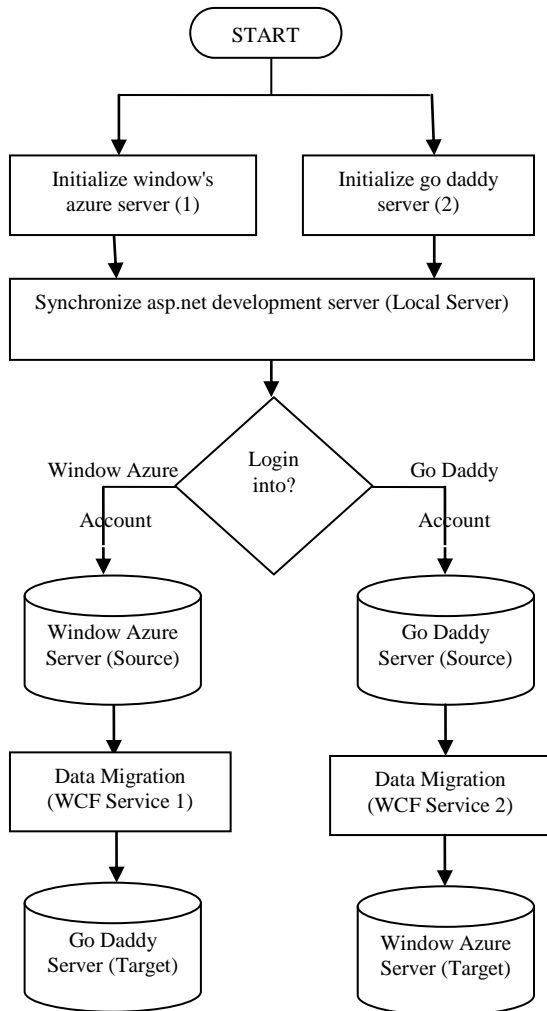- No need to know API of cloud storages to move the data because it uses endpoint service.

Fig. 3. Proposed flow chart.

III. PROPOSED WORK

The main objective of this research work to make a protocol service that makes a possible communicate between two different cloud storages having different cloud vendors & also to migrate data & its architecture from one cloud storage to other.

Fig 3 shows a proposed flow of control of research work:

Step1: Initialize the two cloud storages i.e. windows azure cloud storage & go daddy cloud storage.

Step2: Synchronize these cloud storages with local server of asp.net (called asp.net development server).

Step3: To start the migration process, firstly must log into the window azure server and go daddy server. Then start the application.

Step4:

- If want to migration the data from windows azure to go daddy then logged into the application login and then select data to migrate. WCF Service 1 migrate that data to the go daddy server.
- If want to migration the data from go daddy to window azure then logged into the application login and then select data to migrate. WCF Service 2 migrate that data to the window azure server.

IV. IMPLEMENTATION

A. System Architecture-

Fig 4 shows system architecture of an application and the whole application is divided following modules.

- GUI Module: Every application has a one user interface from where the entire application can be accessed. The user interface of this application is completely based on the end users i.e. it is totally friendly and visually attractive. Here login facility provided to the user to check the authorized user. User filled his details and logged into a system.
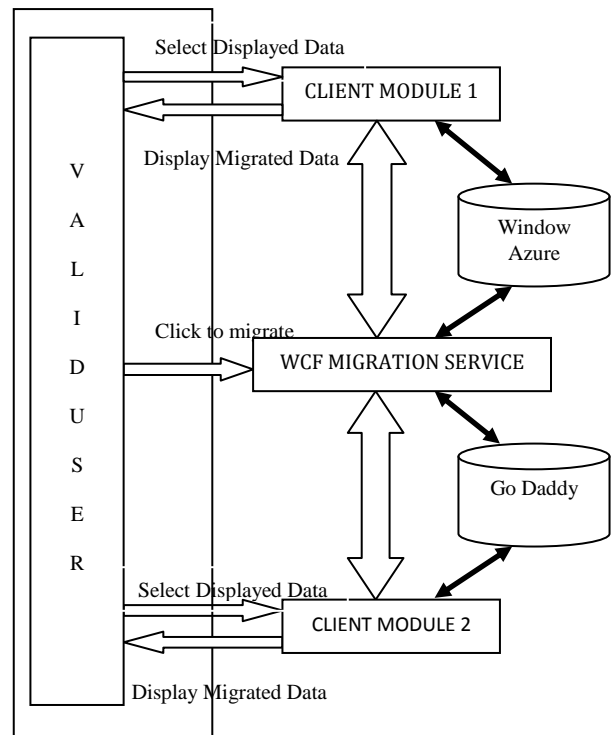

Fig. 4. System architecture.

- Client Module 1: If the data migration process, from windows azure to go daddy then this module displays

all the data that reside on the window azure cloud storage. To migrate the data, user can select the displayed data and click to migrate it to target i.e. to go daddy cloud storage.

- Client Module 2: If the data migration process, from go daddy to windows azure then this module displays all the data that reside on the go daddy cloud storage. To migrate the data, user can select the displayed data and click to migrate it to target i.e. to windows azure cloud storage.

- Migration Module: this module consist the code generated by WCF service that convert the selected data into a common format and then migrate it to the target cloud storage.

B. Application Working-

This whole research work is implemented in VS 2010 Framework with backend of two cloud storages i.e. Window azure and go daddy cloud storage. This Application is divided into two main parts as:

- User Authentication: is a process that checks whether the user is valid or invalid.
  Step 1: Login into the window azure cloud storage. The website URL as:
  http://azure.microsoft.com

  Step 2: Login into the go daddy cloud storage. The website URL as:
  www.godaddy.com
  Step 3:
  a. If user want to migrate the data from window azure to go daddy then login into the application by filling the required credentials and these credentials are matched to correct credentials as stored in window azure cloud storage as shown in Fig 5. Only after successful login, user can take benefits of application.
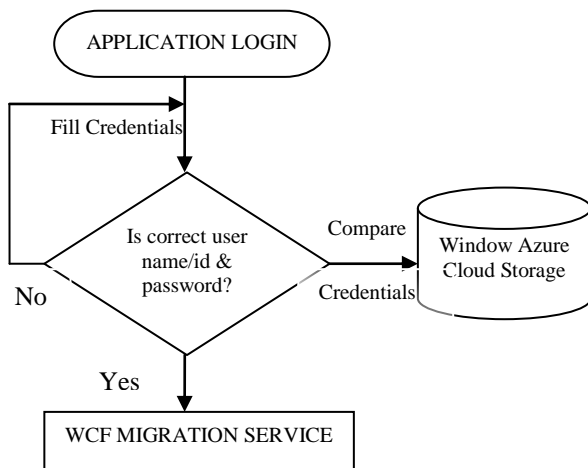

Fig. 5. Application authentication for data migration process from window azure to go daddy.

  b. If user want to migrate the data from go daddy to window azure then login into the application by

filling the required credentials and these credentials are matched to correct credentials as stored in go daddy cloud storage as shown in Fig 6. Only after successful login, user can take benefits of application.
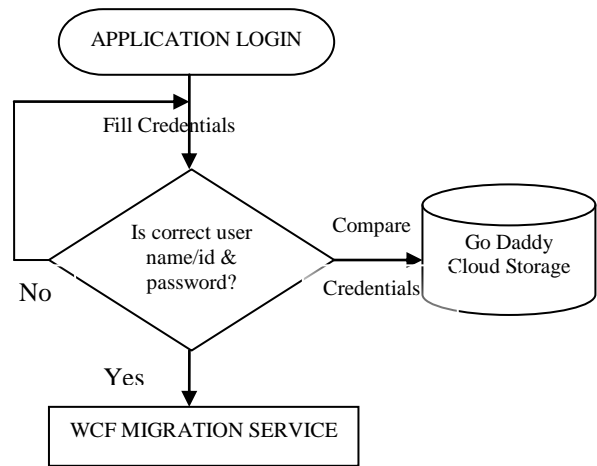

Fig. 6. Application authentication for data migration process from go daddy to window azure.

- User Virtualization: An authorized user can view the various modules of the service as shown in the Fig 7.

  a. Existing scheme module: this module shows the whole data related to login user that present at the source cloud storage.
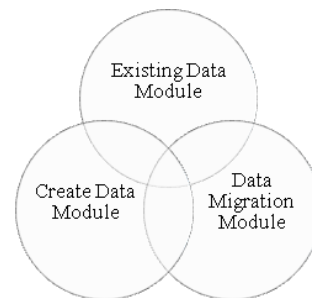

Fig. 7. Application modules.

  b. Creating scheme module: this module creates the data into the source cloud storage.
  c. Data migration module- This module provides the data migration facility to the authorized user from source to target cloud storage

C. WCF Migration Service Algorithm-

- Get information schema and its structure that contains table name, column name, their type, size & its other related schemes like store procedure, view etc from the source cloud storage.
- Manage all columns with related data types, data that reside in tables and in other schemas like store procedure, view etc.
- Convert it into target storage format and store it into the target cloud storage.

D. Security Implementation-

To provide the secure communication to migrate the data, security must be implemented on WCF Migration Service (https) and also on both cloud storages (https) as shown in Fig 8.
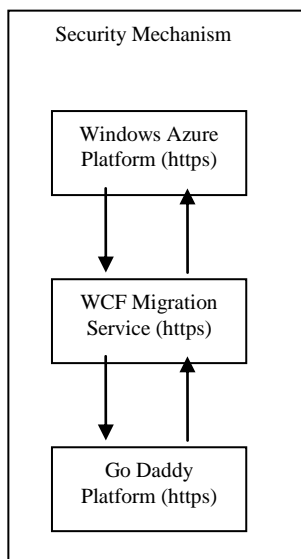
```
┌─────────────────────────────────┐
│      Security Mechanism         │
│                                 │
│   ┌─────────────────────┐       │
│   │   Windows Azure      │      │
│   │   Platform (https)   │      │
│   └─────────────────────┘       │
│        │         ▲              │
│        ▼         │              │
│   ┌─────────────────────┐       │
│   │   WCF Migration      │      │
│   │   Service (https)    │      │
│   └─────────────────────┘       │
│        │         ▲              │
│        ▼         │              │
│   ┌─────────────────────┐       │
│   │   Go Daddy           │      │
│   │   Platform (https)   │      │
│   └─────────────────────┘       │
│                                 │
└─────────────────────────────────┘
```

Fig. 8. Security mechanism.

- There is no need to provide security at window azure cloud storage and go daddy cloud storage because these sites are already support a high level of security (https).
- But the security needed while transferring the data between them so data migration application is built with the transport security to secure the data migration process. Common approach to achieve it is to use secure sockets layer (SSL) for encryption and signing the contents of the data sent over secure Hypertext transfer protocol (HTTPS). It provides a point-to point security between the cloud storages.
- Built application used the IIS and a self created certificate to achieve security.
- After deploying a WCF server it gives the ABC (Address, Binding and Contract) of endpoint. This endpoint helps to uses a particular WCF service. A simple example of ABC of endpoint in WCF as:
  ```
  <client>
  <endpoint address="https://abc-
  pc/WebSite/Service.svc"
  binding="basicHttpBinding"
  bindingConfiguration="secureHttpBinding1"
  contract="ServiceReference1.IService"
  name="secureHttpBinding" />
  </client>
  ```
- Here, ABC of endpoint in WCF
  Address: https://abc-pc/WebSite/Service.svc
  Binding: basicHttpBinding
  Contract: ServiceReference1.Iservice

V. RESULTS

The simulation result of the proposed work is taken out, when three different sizes of data (MB) i.e. 1.04MB, 4.16MB and 8.38MB is migrated from source to target database with internet speed of 100kb/s.

- Fetching Time/Read Performance: Time taken to fetch the records from source cloud storage. Calculated results shown in Table 1, Table 2 & Its graphical representation shown in Fig 9, Fig 10 resp.

Table 1. Fetching time of data migration process from window azure to go daddy cloud storage.

Data (MB)	1.04	4.16	8.34
Fetching Time (hour/min/sec/ms)	00:00:11.2666444	00:02:10.9614906	00:04:47.3127061

Table 2. Fetching time of data migration process from go daddy to window azure cloud storage.

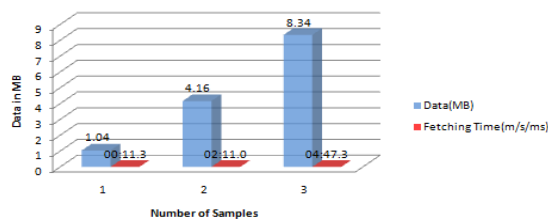| Data (MB)	1.04	4.16	8.34
Fetching Time (hour/min/sec/ms)	00:00:07.1324079	00:00:43.7265010	00:01:49.1032404

Fig .9. Fetching time of data migration process from window azure to go daddy cloud storage.
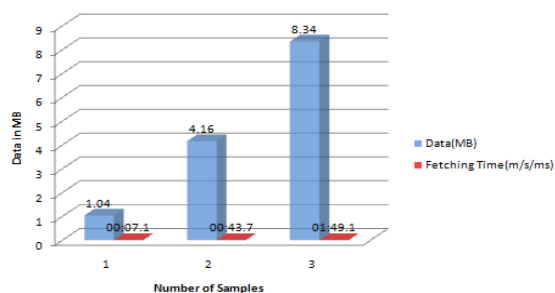

Fig. 10. Fetching time of data migration process from go daddy to window azure cloud storage

- Loading Time/Write Performance: Time taken to load the data into the target cloud storage. Calculated results shown in Table 3, Table 4 & Its graphical representation shown in Fig 11 & Fig 12 resp.

Table 3. Loading time taken during data migration process from window azure to go daddy cloud storage.

Data (MB)	1.04	4.16	8.34
Loading Time (hour/min/sec/ms)	00:02:08.5183505	00:08:39.9557396	00:16:42.9573671

Table 4. Loading time taken during data migration process from go daddy to window azure cloud storage.

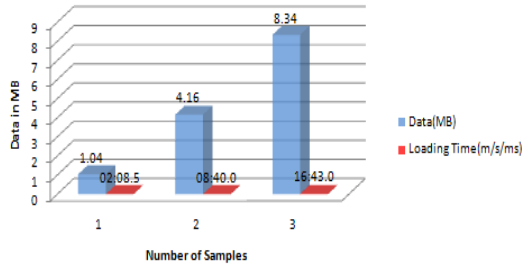| Data (MB)	1.04	4.16	8.34
Fetching Time (hour/min/sec/ms)	00:00:34.6349810	00:02:57.2091360	00:10:37.6634721

Fig. 11. Loading time taken during data migration process from window azure to go daddy cloud storage
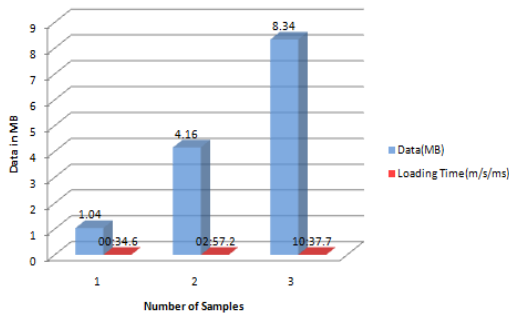

Fig. 12. Loading time of data migration process from go daddy to window azure cloud storage

- Total Migration Time: Time needed to fetch the data from source cloud storage and load it into the target cloud storage.

    Total Migration Time = Fetching Time + Loading Time + Time taken by WCF to manage the data.

    Table 1 shows the results of the data migration process from window azure to go daddy cloud storage. Calculated results shown in Table 5, Table 6 & Its corresponding graphical representation shown in Fig 13, Fig 14.

Table 5. Total Migration time taken by data migration process from window azure to go daddy cloud storage.

Data (MB)	1.04	4.16	8.34
Loading Time (hour/min/sec/ms)	00:02:25.4443189	00:10:53.9654047	00:17:33.3392475

Table 6. Total Migration time taken by data migration process from go daddy to window azure cloud storage.

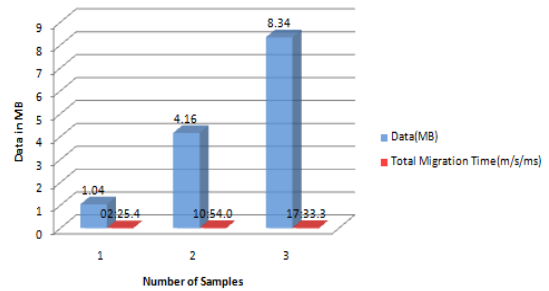| Data (MB)	1.04	4.16	8.34
Loading Time (hour/min/sec/ms)	00:00:46.2066429	00:03:41.8926916	00:12:31.1849654

Fig. 13. Total Migration time taken by data migration process from window azure to go daddy cloud storage
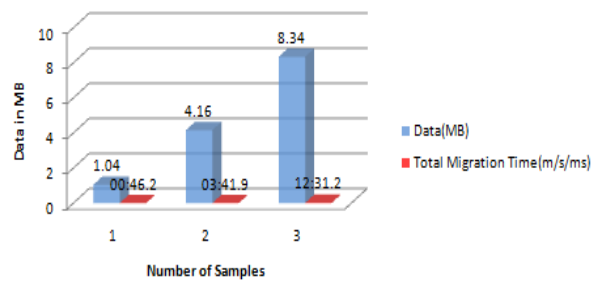

Fig. 14. Total Migration time taken by data migration process from go daddy to window azure cloud storage

- Accuracy = no. of tables + how much data inside it + no. of columns + no. of rows + no. of other scheme and data inside it (e.g. no. of stored procedure and data inside it etc). So, Error rate = 100 − Accuracy. The accuracy calculated after completion of both data migration process is 100% because this work is tested on small size of data i.e. in some MB. There may be an error occur when the data is migrated in large range i.e. in GB or TB.
- Below Table 7 shows that time taken by migration process from window azure to go daddy is more than the data migration process from go daddy to window azure. This means that read performance of go daddy is more than window azure and write performance of window azure is more than go daddy cloud.

Table 7. Read & write performance comparison of window azure & go daddy cloud storage

Data (MB)	Total Migration Time (from window azure to go daddy) (hour/min/sec/ms)	Total Migration Time (from go daddy to window azure) (hour/min/sec/ms)
1.04	00:02:25.4443189	00:00:46.2066429
4.16	00:10:53.9654047	00:03:41.8926916
8.34	00:17:33.3392475	00:12:31.1849654

VI. CONCLUSION & FUTURE WORK

The main vision of this work is to provide an intermediate data migration service called "WCF Migration Service" that supports both heterogeneous & homogeneous cloud storage to achieve the ACID

(Atomicity, Consistency, Isolation and Durability) property of the data.

Future scope:

- Reduce more WCF service migration time & complexity as far as it can be possible.
- This service migrates based on characterize and integer/numeric data It can be extended by transferring the data of other types like images.
- This application can be extended by transferring the data between other different vendor's cloud storages. Only need to know the rules (format they support to store data, what are the credentials to create the data on these, etc) of various cloud storages.

REFERENCES

[1] Abdelsalam Maatuk, Akhtar Ali, and Nick Rossiter, "Relational Database Migration: A Perspective," *Springer-Verlag Berlin Heidelberg- S.S. Bhowmick, J. Küng, and R. Wagner (Eds.): DEXA 2*, pp. 676–683, 2008.

[2] Fábio Luiz Leite Jr., André Gomes de Sousa, et al., "Migratool: Towards a Web-Based Spatial Database Migration Tool*," IEEE Computer Society- Proceedings of the 16th International Workshop on Database and Expert Systems Applications (DEXA'05)*, 2005.

[3] Haji Binali, Chen Wu, and Vidyasagar Potdar, "Web Data Migration: Connecting Databases in the Cloud," *7th IEEE International Conference on Computer Science and Technologies (IEEE DEST 2013)*, 2013.

[4] Kalpana Parsi, and M.Laharika, "A Comparative Study of Different Deployment Models in a Cloud," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, pp. 512–515, May 2013.

[5] Kola Siva Tharun1, Muppalla Prudhvi2, and Satti Swami Reddy3, "ADVANTAGES OF WCF OVER WEB SERVICES," *International Journal of Computer Science and Mobile Computing*, vol. 2, pp. 340–345, April 2013.

[6] Larry Franks, and Sidney Higa, et al., "Window Azure Prescriptive Guidance," *Microsoft Corporation,* 2012.

[7] Nasuni, "Bulk Data Migration in the Cloud," White paper.

[8] Oracle, "Successful Data Migration," White Paper.

[9] Pravin O. Balbudhe, "Cloud Storage Reference Model for Cloud Computing," International *Journal of IT, Engineering and Applied Sciences Research*, vol. 2, pp. 81–85, March 2013.

[10] Shrimant B. Bandgar, Manojkumar Bhosle, T. M. Bansod, and B. B. Meshram, "Design of Distributed System application on SOA, WPF, WF and WCF like new technologies: a Survey," *International Journal of Management, IT and Engineering*, vol. 2, pp. 319–329, June 2012.

[11] Tharam Dillon, Chen Wu, and Elizabeth Chang, "Cloud Computing: Issues and Challenges," *IEEE Computer Society- 24th IEEE International Conference on Advanced Information Networking and Applications*, 2010.

Authors' Profiles

Parvinder Kaur: I did my B. Tech (Information Technology) from Adesh College of Engineering and Technology, Faridkot, Punjab, India in 2012. Currently, I am pursuing M. Tech (Information Tehnology) from Chandigarh Engineering College, Landran, and Punjab, India. My interest is in Web Technologies and Web Mining.

Manish Mahajan: He is working as an Associate professor in the Department of Information Technology, Chandigarh Engineering College, Landran, and Punjab India. He has completed his M-Tech in CSE from PTU and also pursuing Ph.D. from Punjab Technical University, Punjab. His research interests include Digital Image Processing, Data Mining, Steganography and Data Security.